Invention Journal of Research Technology in Engineering & Management (IJRTEM)
ISSN: 2455-3689
www.ijrtem.com Volume 2 Issue 4 ǁ April 2018 ǁ PP 48-55

Modeling the Student Success or Failure in Engineering at VUT Using the Date Band Algorithm

[1,]Langa Hendrick Musawenkosi , [2,]Twala Bhekisipho[2]
[1]*Department of Electrical Engineering, Vaal University of Technology, Vanderbijlpark, South Africa*
[2]*Department of Electrical & Mining Engineering, University of South Africa , Florida, Johannesburg, South Africa*

ABSTRACT: *The success or failure of students is a concern for every academic institution, college, university, governments and students themselves. This paper presents a model to determine the propensity of a student to succeed in the Electrical Engineering Department at Vaal University of Technology. Firstly, various machine learning algorithms which can be used in modelling and in predicting student success or failure are discussed as well a new algorithm called the date band algorithm. Secondly, the concept of an academic model is also discussed. This model defines the domain and focus of data used to make preditcions. The academic model consists of the subject, the lecturer and the sudent each of which has various attributes. One of the attributes discussed in this paper is the popularity index which is a measure of cohesiveness of the model.The Date Band Algorithm is presented among others in the development of the model. In this algorithm, predictions are made to optimize the performance of academic environment, thereby impacting on the choices of funders when they support students.*

KEYWORDS: *Academic Environment Model, Date Band Algorithm, Decision Trees, K-Nearest Neighbor, Machine Learning;*

I. INTRODUCTION

The success rate in academic environments is not only a concern for those institutions but also governments, sponsors such as the public and the private sectors, parents, students themselves and other stakeholders. It therefore behooves us to investigate the propensity of those students to succeed using scientific methods such as machine learning. Machine Learning (ML) has a variety of algorithms that can be applied in addressing this problem.The South African government and funders can save a lot of resources when funding these institutions. Therefore, the application of rigorous methods of machine learning can improve the efficiency in the academic sector. For the most part in South Africa, the largest contributor of funding in public education is government, that is, the ministry of education. Although the ministry of education takes no account of income that is raised from student fees and other private sources, these public institutions have to account by submitting annual financial statements which reflect all income and all expenditure from all public and private sources [8].

The need to attract and retain students in engineering programs remains by necessity a focal point of interest and effort in engineering education, [12]. All universities and colleges have marketing departments to make sure that they attract the best of the best. They run various marketing programs for this purpose. Paul and Cowe Falls, [3] highlighted the three aspects for engineering careers success based on the availability of the resources. Firstly, lifelong learning is fundamental for success in the 21st century engineering career. Staying abreast with the most recent technological advancement is essential for being innovative and creative. Secondly a study in the engineering construction industry is the most critical aspect of fostering a successful career path was in developing a career network. This includes networking, mentorship training and constructive feedback. Thirdly, the aspect of engineering career success relates to the models "proactive personality" variable.

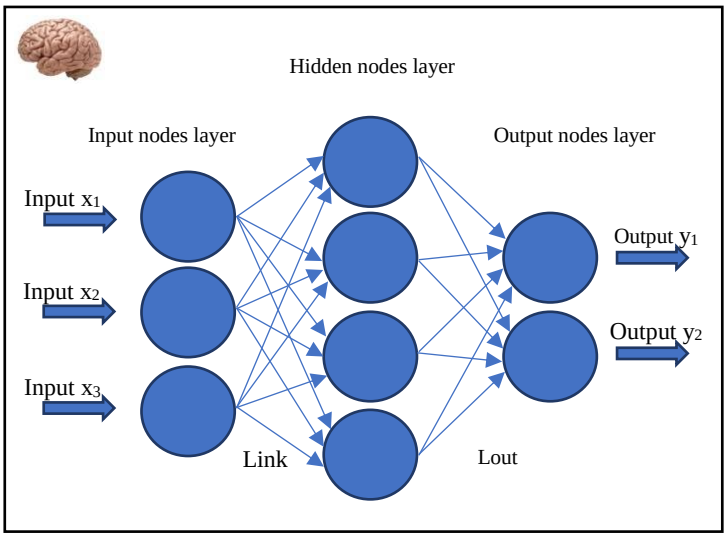
II. MACHINE LEARNING ALGORITHMS OVERVIEW

Decision Trees, DTs : Decision Trees, DTs [12] are simple yet successful techniques for supervised classification learning. This classification method consists of decision nodes, connected by branches, extending from the root node until the terminating leaf nodes, [1]. Starting at the root node attributes are tested at the decision node, with each possible outcome resulting in a branch. A decision tree algorithm aims to recursively split the observations into mutually exclusive subgroups until there is not further split that makes a difference in terms of statistical or impurity measures, [4]. A path is traced from the root to a leaf node which holds the class predication for that sample. Decision trees can easily be converted into IF-THEN rules and used for decision making, [9].

The K-Nearest Neighbor Classifier, KNN : The K – Nearest Neighbor is an example of instance based learning, in which the data set is stored, so that the classification for a new unclassified record may be found by simply comparing it to the most similar records in the training set [5]. The KNN algorithm is the earliest researched algorithm used for classification and is proved as one of the algorithms which have good classification results, but there are still some problems that need to be attended to. For example, it is not yet settled, how to select the value of k and how to select feature sets to make the classification better and their impact on each other [11].

The Support Vector Machine, SVM : Support Vector Machines (SVM) is an algorithm that uses nonlinear mapping to transform the original data into a higher dimension [7]. SVM's are pattern classifiers that can be expressed in the form of hyper planes to discriminate between positive instances and negative instances pioneered by Vapkin [10].

The Artificial Neural Network, ANN : The artificial neural network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, like the brain process information [6]. ANN's are powerful tools that can be used to learn patterns from data. According to [5] a neural network is a collection of nodes that are connected in some pattern to allow communication between the nodes. These nodes, also referred to as neurons or units, are simple processors whose computing ability is typically restricted to a rule for combining input signals and an activation rule that takes the combined input to calculate the output signal. Output signals may be sent to other nodes along connections known as weights. The weights usually excite or inhibit the signal that is being communicated.



An example of neural network is shown in Fig. 1. Since the artificial neuron mimics the natural neuron, neural networks therefore mimics the operation of the brain and can be quite useful in making classifications and predictions.

III. MODELING AN ACADEMIC ENVIRONMENT

The Academic Environment Model : The environment is the sum total of surroundings of a living organism including natural forces and other things, which provide conditions for development and growth as well as danger and damage. An academic environment, in figure 2, is where a student exists. In order to model this, it is necessary to gather information about the student, the lecturer and the module which will form part of the environment.

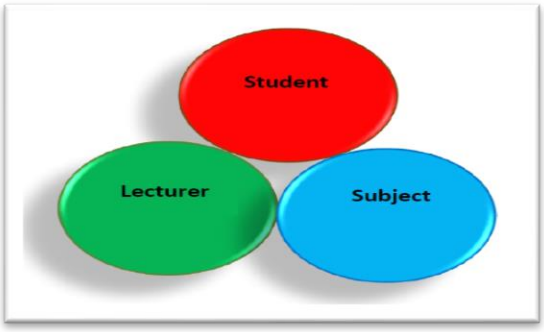


Fig. 2. Academic Environment Model

Tinto's (1975) Student Integration Model (SIM) postulated that students who persist and succeed in college are those who are able to integrate successfully into an institution's social an academic environment. Alternatively, the students who are more likely to struggle and fail to persist are those who do not attempt or achieve social and academic integration.

The Lecturer's Popularity Index : One of the common problems in higher education is the evaluation of the instructor's performances in a course. The most widely used tool to evaluate instructors' performance in a course is through surveying students' responses about the course and its instructor through a questionnaire [2]. The percentage score of likes for a given the subject or the lecturer is given by the following equation:

$$L(x) = \frac{\sum_{i=1}^{n} x_i}{nk} \qquad (1)$$

Where:
n = number of instances of likes for a lecturer or subject and
k = the highest number of likes in an instance
Clearly, some lecturers are more popular than others. There are lecturers whom students really detest and there are lecturers whom they adore. These can be due to several reasons, such as the appearance, teaching style, level of education, leadership and so on. The popularity of the lecturer has a correlation with the performance of the student. Similarly, the popularity of the subject can be measured.

The Academic Environment Client System : The panel below allows the student to rate the lecturer using choices of numbers between 1 and 5 as shown in Fig 3. If the lecturer is least popular then the choice would be a 1 and if the lecturer is a student's favorite then the choice would be a 5. This information is then captured in a database for future references. The popularity of the lecturer can increase or decrease with time depending on the performance of the lecturer. This is an important feature to have in the design.

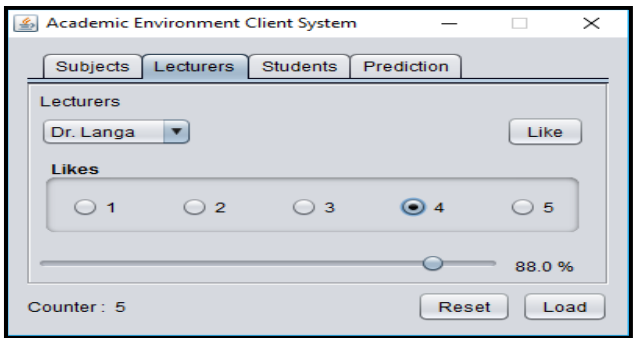


Fig. 3. Academic Environment Client System

A survey of approximately 600 students was completed where eleven lecturers were evaluated and eight subjects were also evaluated using the system as shown in the figure below. Upon analyzing results, interesting observations were noted. A lecturer teaches an average of 120 students per semester.

Popularity Indices for Power Engineering Lecturers : It is clear from the graph that some lecturers are more popular than others and that no lecturer obtained less than 50% which is fair enough in terms of the quality of lecturers employed in the department of Power Engineering. There are various reasons why a lecturer would be unpopular. It could just be shear laziness on his part, lack of understanding of the subject he teaches, the attitude, very strict, to name a few but a few. And there are various reasons why a lecturer could be popular. It could be that they are good in the subject matter, they have good qualifications, they are lenient, their attitude, again to name a few.

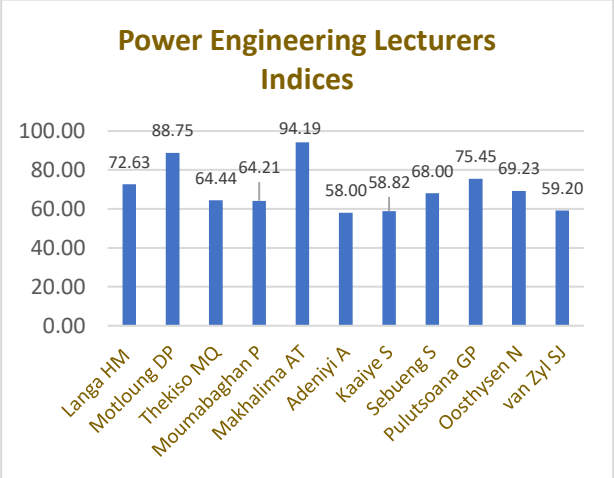


Fig. 4. Popularity Indices for Power Engineering Lecturers

Evidently, as seen in Fig. 4, some lecturers are more popular than others in this department as would be the case with other departments.

Throughputs for 2016 semester 1 & 2 : There are many ML algorithms used for making predictions. This paper focuses on the algorithm called the date band algorithm with special reference to making predictions of student success or failure. Some algorithms will be more accurate than others and some will be more appropriate than others also. Evidently, as shown in Fig. 5, some subjects are more popular than others in this department as would be the case with other departments.

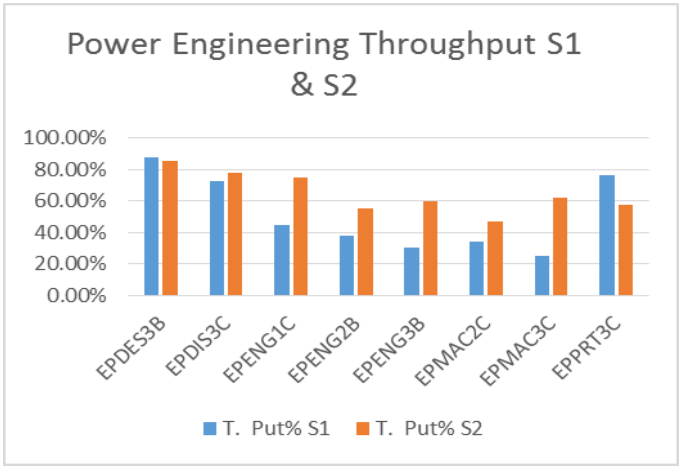


Fig. 5. Throughput for Semester 1 & 2 in 2016

The x axis is the representation of subject codes for the department of Power Engineering. T Put% S1 and T Put% S1 are percentage throughputs for semester 1 and 2 respectively. An interesting observation in the graph is that there is a general improvement in the throughput for all subjects except for EPPRT3C and a slight decrease in EPDES3B. While it is plausible to observe an increase in the throughput it is conversely concerning to see a decline in the throughput. This decline in the throughput, however, can be addressed with the lecturers who are responsible for teaching those subjects. On the other hand, for those subjects that have seen improvement, either maintain the position or strive for more improvement.

IV. THE DATE BAND ALGORITHM

The date band algorithm, shown in Algorithm 1.1, is derived from an assumption that there exists a relationship between the date of birth of the student attribute and the propensity of that student to pass or fail. The algorithm scans the date band to determine the number of students that succeed in that band and those that fail in that band. To predict the probability of the student instance to succeed, it is a question of classifying the student according to that probability.

Algorithm 1.1 Date Band Algorithm

```
Start
{
int Count = 0, Passed = 0, n = 0;
Enter date of birth
Read date of birth,
for ( n = 1, to number of entries, n++)
{
Scan the list of students for students within this date band
If (Found) then
{
Count++;
Check if Successful
If (Successful) then
{
Passed++
}
End if
}
End if
}
End for
Predict Probability for Success
}
End Start
```

So, this algorithm, as it has been stated, works on the premise that students can first be classified in date bands as far as their date of birth are concerned and then a prediction of the probability of that student to succeed or fail in the academic environment can be made. The following graph reveals some interesting results.

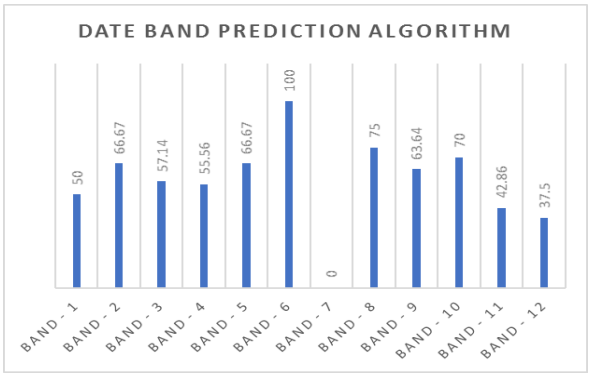


Fig. 6. Date Band AlgorithmProbabilities

Fig. 7.

It is clear from this bar graph, in Fig. 6, that Date band – 6 scores the highest, and this means that the students in this band have 100 % probability that they will succeed in when enrolled for this subject. The lowest in this test is Band – 12 with 37.5 % probability. Band – 7 had a 0 % probability, but the reason it is zero is that there were no students in that band in fact.

V. RESULTS OF THE DATE BAND ALGORITHM

The academic environment consists of the student, the lecturer and the course or subject that the student enrolls for. It has been shown that the condition of the environment can be measured by making use of the popularity indices of lecturers and the courses themselves. So, when a student enters an environment, he or she will be subjected to these conditions. On the other hand, the student has certain features which play a part in determining his or her success. These features can include the APS score. The target variable could pass or fail, which implies that we want to predict whether the student will pass or not given certain set of variables such as:

X_1 = English
X_2 = Mathematics
X_3 = Physical Science
X_4 = Students Birth Date Band

These variables were used also in the multiple regression model and the decision tree model to realize similar results as shown in Fig. 7.

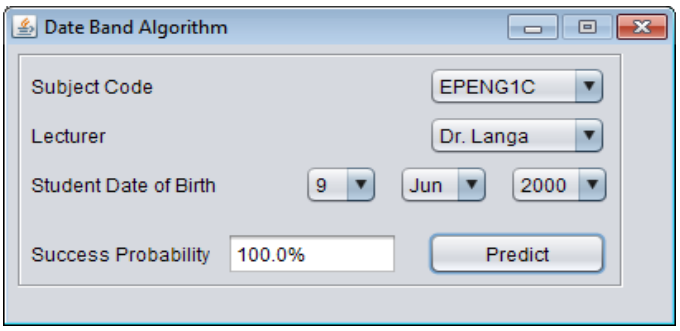


Fig. 8. Date Band Algorithm GUI

Other variables or features can be considered such a as age, place of birth, household income etc. and they would have varying degrees of influence in the prediction outcome. For example, in the KNN algorithm, it has been found that the most influential variables are Mathematics, Physics and Chemistry and English is least influential. In our algorithm, we use the Date Band as an input variable although a combination with other features can be performed as in a neural network for instance. A common standard of entry requirements for engineering schools in South Africa are Mathematics, Science and English.

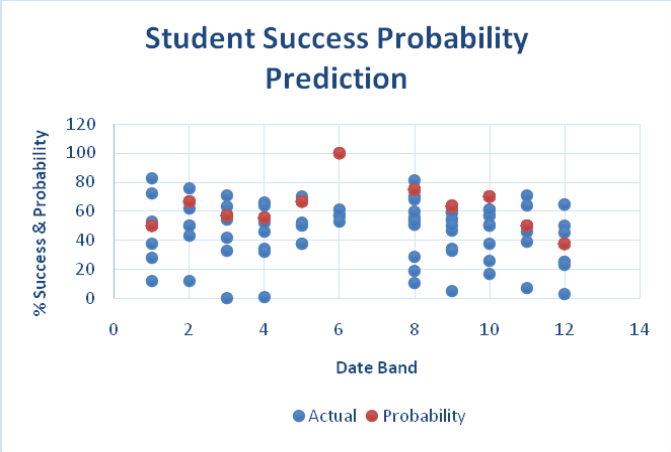


Fig. 9. Student Success Probalitiy Prediction Model Rerults

It is interesting in Fig. 8, to note that there is no data available for Band 7, therefore a prediction cannot be made. This is because of the lack of historical data with which to make predictions. It is also that for Band 6 there is a 100% chance for success but it does not mean that, that position will necessarily remain the case because input data sets change with time. When using the Date Band Algorithm, it has been found that a prediction can be made on the basis of the environmental conditions that, it is known how the class of students are doing given a lecturer and a subject and the most recent performance of that class. Consequently, it becomes a very dynamic model.

VI. CONCLUSION

This paper has presented the Date Band Algorithm as one of the solutions to predict the probability of a student who wishes to enroll for a subject to succeed. So, the basic question that is asked by a student who wishes to take a subject is this, "What are my chances of succeeding should I take a particular subject with lecturer so and so?" He then enters that information in the system and the algorithm will generate a probability which can be essentially regarded as an advice, a warning or information. It is in this paper, finally that the Date Band algorithm was used to predict the propensity of the student who enters the academic environment in the department of Power Engineering at the Vaal University of Technology.

REFERENCES

1. Larose, T.D., Larose, D.C.: Data Mining and Predictive Analytics, 2nd edn, Wiley, (2015).
2. Agaolgu, M.: Predicting Instructor Performance Using Data Mining Techniques in Higher Education, IEEE, pp. 2379 – 2387, Turkey (2016).
3. Paul R.,Cowe Falls, L.: Mapping Career Success Competencies to Engieering Leadership Capabilities, IEEE, pp. 1 – 6, Galgary, Canada (2015).
4. Agaolgu, M.: Predicting Instructor Performance Using Data Mining Techniques in Higher Education, IEEE (2016).
5. Callan, R: Artificial Intelligence, Palgrave Macmillan (2003).
6. Abdella, M., Marwala, T.: Treatment of missing data using neural networks and genetic algorithms. Proceedings of the International Joint Conference on Neural Networks, July 31 – August 23, pp. 598 – 603, Montreal, Canada (2005).
7. Han, J., Kamber, M., Pei, J.: Data Mining 3rd edn, Morgan Kaufmann Publications (2012).
8. Ministry of Education. A New Funding Framework: How Government grants are allocated to Higher Education Public Institutions. (February 2004).
9. Kamber, M., Winstone, L., Gong, W., Cheng, S., Han, J.: Generalization and Decision Tree Induction: Efficient Classification in Data Mining, IEEE, pp. 111 – 120, (1997).
10. Twala, B.: Robot Execution Failure Prediction Using Incomplete Data. Proceedings of the 2009 IEEE International Conference on Robotics and Biometrics, December pp. 1518 – 1523, China (2009).

11. Shang, W., Zhu, H.: The Improved Ontology kNN Algorithm and its Application, pp. 198 – 203, IEEE, (2006).
12. Imbrie, P.K., Lin, J.: Work in Progress Engineering Students Change in Profile Over the Freshman Year across Male and Female Samples: A Neural Network Approach, 36th ASEE / IEEE Frontiers in Education Conference, pp. 27 – 28, San Diego, CA (2006).